REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
03/25/2016	Final Report	7 Feb 2014 - 25 Mar 2016

4. TITLE AND SUBTITLE
Early Detection of Risk Taking in Groups and Individuals

5a. CONTRACT NUMBER
N00014-14-C-0047

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
MacGregor, Donald G.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
MacGregor Bates, Inc.
1010 Villard Avenue
PO Box 276
Cottage Grove, Oregon, USA 97424

8. PERFORMING ORGANIZATION REPORT NUMBER
MBI-2016-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 North Randolph St., Suite 1143E
Code: 30
Arlington, Virginia 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited distribution

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Previous research in behavioral psychology offers potential methods for analyzing social media that can indicate the presence of conditions conducive to risk-seeking. This project applies such methods to Twitter messages from a range of social events, some where disruption is present and others where it is not. The methodologies focus on message content and network connectedness. The results indicate that conditions conducive to risk-seeking are present in some social events, as assessed by linguistic markers that reflect the relative balance of reasoning and emotion, as well as emotion-specific conditions relating to anger.

15. SUBJECT TERMS
social media analysis, risk-taking, group polarization

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Donald G. MacGregor, Ph.D.
U	U	U	SAR	39	19b. TELEPHONE NUMBER *(Include area code)* (541) 942-5727

FINAL REPORT OF PROJECT

"Early Detection of Risk Taking in Groups and Individuals"

OFFICE OF NAVAL RESEARCH

(Contract No. N00014-14-C-0047)

Principal Investigator:

Donald G. MacGregor, Ph.D.[1]

ABSTRACT

Previous research in behavioral psychology pertaining to risk-related behavior offers potential methods for analyzing social media that can provide an indication of the presence of conditions conducive to risk-seeking. This project applies such methods to Twitter messages from a range of social events, some where disruption is present and others where it is not. The methodologies focus on two network characteristics: message content and network connectedness. The results indicate that conditions conducive to risk-seeking are present in some social events, as assessed by linguistic markers that reflect the relative balance of reasoning and emotion, as well as emotion-specific conditions relating to anger. In addition, a word usage category, *Redemption*, derived from Prospect Theory (Kahneman & Tversky) and that represented a call to return social conditions to a more desirable and just state related both to emotion-related word categories as well as network connectedness. The combination of standard word categories from LIWC, augmented by an additional behaviorally-derived linguistic marker (*Redemption*), and network connectedness metrics provided a basis for differentiating between social events.

[1] MacGregor Bates, Inc., 1010 Villard Avenue, P.O. Box 276, Cottage Grove, OR, 97424, USA. Tel: 541-942-5727 FAX: 541-942-8041 Electronic: donaldm@epud.net.

Introduction and Background

Risk-taking behavior has long been a topic of interest in the social and behavioral sciences. Much of the research in this area has been done in the context of health and safety, with an interest in understanding the underlying factors that promote risky behaviors such as tobacco, alcohol and substance abuse particularly in young people. The importance of examining linguistic discourse from the perspective of risk-taking lies in the fundamental nature of human behavior as goal driven, and in the nature of extreme behavior (e.g., aggression, violence) to place an individual or a group at-risk when their actions have the potential to have undesirable consequences to themselves. This is the case when individuals in groups engage in social protests or demonstrations that place them in situations where police or other authorities are confronted and their behavior constitutes risk-taking.

Research in the psychology of risk has identified a number of factors that are associated with how people evaluated situations with respect to risk and how risk-taking enters into reactions to social conditions.[2] One of the factors that has been demonstrated to be significant in risk-taking attitudes is perception of control, where a greater degree of perceived control over the consequences risk-taking behavior tends to be associated with more extreme risk taking behavior. A consistent finding in the risk literature is a generalized *optimism bias*: across a range of risk-taking contexts (e.g., driving, tobacco use) the tendency is to believe negative outcomes are less likely to happen to oneself than they are to others.[3] We discuss these and other risk-related concepts in the sections below as they apply to the analysis of social media.

In all of these contexts the methods of study have generally used approaches such as group/individual interviews and survey-based methods. Although these approaches have provided much in the way of theoretical development and empirical support for interventions, they rely on the availability of groups and individuals for study. A more significant problem is posed when the groups and individuals of interest are not as directly accessible to the application of such methods, as is the case in many non-western cultures where access to general populations is limited or restricted.

A pathway forward to resolving this dilemma is offered by new capabilities for both capturing and analyzing social media using theoretical approaches that provide linkages between social (group) organization, group cohesion and risk-related behavior. The proposed effort intends to exploit these opportunities by bringing to the analysis of social media theories and concepts from the psychology of risk and behavioral decision theory that point to the potential utility of linguistic markers as early detectors of risk-taking behavior.

[2] e.g., Slovic, P. (2000). *The perception of risk.* London, UK: Earthscan Publications Ltd.

[3] e.g., Slovic, P., Fischhoff, B., & Lichtenstein, S. (1978). Accident probabilities and seat belt usage: A psychological perspective. *Accident Analysis and Prevention*, 10, 281-285.

Our working hypothesis is that research in the psychology of risk, behavioral decision theory and linguistic analysis can be used to mutually inform and guide the development of methods of social media analysis. This hypothesis leads us to draw upon a number of theories and empirical findings from the psychology of risk as well as behavioral decision theory to identify linguistics markers that pertain to risk-taking and that can be used as a basis for the analysis of social media. The technical approach is based on the concept that multiple risk-related factors and propensities interact to produce the potential for risk-taking, and that these risk-related factors are detectable in the linguistic discourse that is contained in microblogs. The risk-taking behavior that is the ultimate focus of interest is that associated with disruptive, aggressive, unlawful, and (potentially) destructive activities that pose a hazard to civilian and military populations.

The data that serves these research purposes has come from real-time social media in the form of Twitter messages created during differing social events that vary in terms of social disruption and violence toward authority figures. The purpose of the research is to identify the presence of conditions that are conducive to risk-taking behavior. In pursuit of this objective, we analyze data from social events where such conditions are present, as well as social event where they are not. The primary focus is on linguistic analysis that seeks to identify psychological concepts that research has shown have a relationship to risk seeking. We also explored network characteristics that relate to group identity and cohesion, which also have potential for risk-taking behavior.

Network Characteristics of Twitter

Twitter was launched on the Internet in 2006 (www.twitter.com) as a unique opportunity for people to simply post a brief (140 character maximum) message that was essentially an answer to the question "What are you doing right now?" Twitter messages, generally referred to as "tweets" have over the past six years grown substantially in volume both in the U.S. and in other countries. Users can peruse ongoing Twitter conversations and enter the discussion at will. Because of the 140-character constraint on tweets, messages may contain short phrases or truncated expressions rather than in-depth discussions such as might be found on regular blogs, discussion boards and the like. Ye & Wu (2010)[4] examined the propagation and network characteristics of tweets and found that messages traveled relatively far, with over 1/3 spreading more than three hops away from the message originator. Moreover, message replies were quite rapid, with 25% of replies within only slightly more than a minute, and 75% within 16 minutes. Conversation flows generally lasted less than an hour and about 25% less than two minutes. Thus, Twitter messages are relatively high volume, receive quick replies and are propagated a relatively far distance from their original source, though

[4] Ye, S., & Wu, F. (2010). Measuring message propagation and social influence on twitter.com. *Proceedings of the 2nd International Conference on Social Informatics* (pp 216-231).

conversations may last for a relatively short period of time. Yang & Counts (2010)[5] studied the propagation of @*username* mentions in tweets, an indicator of conversations, and found that for interaction networks constructed from such messages speed of diffusion (how quickly a tweet will produce a responding tweet) was predictable from the frequency of @*username* mentions. In addition, @*username* mentions was a strong predictor of the number of hops produced by an originating message.

Twitter as Predictor of Social Events

One foci of research on Twitter has been the potential value of tweets as predictors of social events and phenomena. Asur & Huberman[6] developed a predictive model using Twitter "chatter" to forecast box-office revenues for movies. Using a set of 24 movies released in 2009, they examined a database of 2.89 million tweets from 1.2 million users. Using a predictive model based on the *average tweet rate* during the pre-release period, they were able to obtain a linear correlation coefficient in the .90 range between twitter volume and opening weekend box office sales. Expanding their predictive model to include sentiment analysis of tweet content (e.g., positive/negative content) added significantly to their predictive model, but not as substantially as tweet volume.

In a study with similar objectives, Bollen, Mao & Zeng[7] applied a sentimental analysis to twitter content in the form of mood-related variables to predict stock market values (Dow Jones Industrial Average or DJIA). Their method used a battery of emotion-based approaches for codifying the public mood associated with tweet text content during a training and calibration period as the basis for predicting the DJIA during a performance period. Their results yielded a high level of predictability (i.e., $R^2 > .68$) of stock market variance from public mood states.

Twitter has also been used as the basis for prediction in a political context. Tumasjen, Sprenger, Sandner & Welpe[8] used Twitter messages to predict political opinions associated with the German federal parliamentary elections in 2009. They examined three features of political opinion: (a) the role of Twitter in political discussion and deliberation, (b) electorate sentiment about the election as expressed on Twitter, and (c) predictability of election results from Twitter messages. Using a database of approximately 100,000 political tweets, their study extracted sentiment using LIWC2007 (Linguistic Inquiry & Word Count)[9], the study

[5] Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 355-358).

[6] Asur, D., & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. (pp 492-499).

[7] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2, 1-8.

[8] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.* (pp. 175-185).

[9] Pennebaker, J., Chung, C., & Ireland, M. (2007). The development and psychometric properties of LIWC2007. Austin, TX.

found that Twitter was an active platform for political discussion and deliberation. Moreover, Twitter content accurately reflected voter preferences as reflected in poll results. Sentiment analysis indicated that Twitter messages correspond closely in tone to similarly codified results from news media.

In a study that extended the work of Tumasjen, *et. al.*, to the context of the Irish General Election in 2011, Bermingham & Smeaton[10] applied sentiment analysis based upon tweets coding that used a learning/training approach that was initiated by initial classifications of politically-specific sentiment. They found that preference votes in the election were predicted from the share of tweets volume each party received over the study period. Sentiment contributed to the overall predictive accuracy of the model, but presented a more subtle and complex relationship to election outcomes, with sentiment results having the largest impact in a fairly narrow time period around the election day. This suggests that sentiment may play a more volatile role in formulating attitudes and opinions relating political behavior, and could be a better predictor for near-term political behavior than might be gauged from volume of discussion.

Twitter as Predictor of Unlawful Behavior

In a slightly different vein, Twitter has been used effectively to predict unlawful behavior. Wang, Gerber & Brown[11] developed a Twitter-based prediction model that relies on a deeper semantic understanding of tweets than that obtained through tweet volume and sentiment. Linguistic processing based on the use of semantic role labeling was applied (through latent semantic processing) to index key topics relating to criminal behavior. The resulting extracted topics provided additional contextual information. The predictive model was then based on the extracted latent topics. Essentially, their work utilizes conventional semantic labeling to reduce the possible dimensions contained in natural language messages. These reduced dimensions then allow a more meaningful interpretation of message content. This approach offers a great deal of promise in contexts where particular topics of conversation can be identified in Twitter content, and thereby used to refine the codification, labeling and interpretation of conversation.

What distinguishes the work of Wang, et. al. is its use of natural language processing to develop a characterization of a risk-related environment from which behavioral predictions are made. In their case, the environment was that of roadway hazards which have the potential to lead to motor vehicle accidents. The associated unlawful behavior was that of "hit and run" – cases where motorists experience an accident but do not remain at the scene as required by law.

---

[10] Bermingham, A., & Smeaton, A. F. (2011). On using Twitter to monitor political sentiment and predict election results. *Sentiment Analysis Where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP), 13th November 2011, Chiang Mai, Thailand.*

[11] Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extract from Twitter posts. In S. J. Yang, A. Greenberg, & M. Endsley (Eds*). Social computing, behavioral-cultural modeling and prediction*. Berlin: Springer.

Essentially, this approach to processing Twitter messages provides an indication of *opportunities* for risk-related behavior relating to motor vehicle operation.

An extension of this concept can be applied to analyzing Twitter databases for indicators of risk-taking opportunities. In the context of social and political unrest or demonstration, for example, these opportunities could be expressed in tweets as statements or utterances regarding the environment. These could take the form of information regarding the presence or absence of, for example, legal authorities, police, security forces and the like. They could also take the form of location information more suitable to, for example, political or social demonstrations, or where like-minded individuals might congregate or gather. The potential for social dynamics to contribute to risk-taking behavior is discussed in the sections below.

Twitter and Social Dynamics

Social dynamics are a critical element in detecting the presence of conditions conducive to risk-taking behavior. Twitter has been studied in terms of its properties as an indicator of social dynamics, and particularly the information that can be gleaned from the analysis of tweets concerning the phenomenon of *group polarization*. Group polarization refers to a tendency for groups to hold more extreme positions or attitudes than the average of the pre-group attitudes of its individual members. In cases where the attitudes of individuals already tend toward cautiousness, group polarization can tend to amplify that caution. On the other hand, when attitudes tend toward less cautious or risky, polarization can lead a group toward greater risk. The phenomenon of group polarization is important because it helps explain human behavior in a of real-life situations, such as policy decisions.[12] Moreover, these effects have been observed and demonstrated in cultures around the world with varying types of group participants.[13]

Key to group polarization is the interaction of its members in some form of discourse.[14] In the context of computer-mediated communication Sia, Tan & Wei found that even in contexts where only text interaction occurred (i.e., no visual cues) group polarization was not only present, but even more extreme than in cases where the bandwith of interaction was greater (i.e., visual cues).[15] Some commentators have noted that the combination of modern media and the Internet have exacerbated a trend toward polarization that has become more extreme simply because individuals can seek out the attitudes of others with similar views, thereby increasing their confidence in their own opinions.[16]

---

[12] e.g., Whyte, G., & Levi, A. S. (1994). The origins and function of the reference point in risky group decision making. The case of the Cuban missile crisis. *Journal of Behavioral Decision Making*, 7, 243-260.

[13] e.g., Forsyth, D. R. (1990). *Group dynamics*. Pacific Grove, CA: Brooks Cole Publishing.

[14] Van Swol, L. M. (2009). Extreme members and group polarization. *Social Influence*, 4(3), 185-199.

[15] Sia, C. L., Tan, B., & Wei, K. K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence and anonymity. *Information Systems Research*, 13(1), 70-90.

[16] Sunstein, C. (2008). The law of group polarization. In J. S. Fishkin & P. Laslett (Eds.), *Debating deliberative democracy*. Blackwell Publishing, Ltd.

Taken together, these research findings indicate that group polarization can occur without physical engagement, and can be even more potent as a basis for group formation in situations where individuals can self-select themselves anonymously into groups based upon their own prior attitudes and positions. Indeed, the picture that emerges with respect to microblogs is that Twitter can readily foster and sustain the growth of group polarization through the consistent dialogue that is part of "tweeting", no matter what their geographic location.[17].

[17] Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, 30(5) 316-327.

METHODOLOGY

The general model for the study is based on extracting from Twitter messages two basic elements. One is the underlying network that represents the interrelationships between Twitter messages in terms their connectedness to each other. Analyses of these interrelationships provides an indication of the social dynamics present in a message dataset, as measured by network metrics that capture the presence of groupings messages that have implications for social organization.

The second extraction is the content of the messages themselves that form a basis for linguistic analyses to assess the degree to which conditions conductive to risk-related behavior involving violence or other confrontational actions may be present. These analyses are done along lines guided by psychological research that suggests cognitive and emotional conditions that have the potential to associate with risk-related behavior.

The analyses resulting from these two extractions lead to the identification of conditions are conducive to risky social behavior, and that are the result of both social dynamics (as observed in network metrics) and affective conditions (as observed in linguistic analyses).

Linguistic Inquiry and Word Count (LIWC)

A widely-used method for social and psychological analysis of text is the program Linguistic Inquiry and Word Count (LIWC).[18] The rationale underlying LIWC is that language is the carrier of people's thought and emotions and analysis of linguistic expression reveals their mental and emotional states. As such, LIWC is grounded in psychological theory and has been used in numerous studies along these lines to reveal important features of what people are thinking and feeling in a given context.[19] LIWC has been used in a number of studies of social media.[20]

LIWC is based on the empirical finding that spoken and written language reflects important emotional and cognitive states of individuals, and that analysis of such language can reveal important features of these states. The program operates on the basis of a psychometrically constructed and validated set of word categories associated with a larger dictionary. The dictionary contains numerous word categories, with each category comprised of a set of distinct words. For example, the word category *Anger* contains 184 words including "fuming" and "insult*". The wildcard "*" is appended to some words so that multiple variations can be captured

[18] Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates.

[19] e.g., Pennebaker, J. W., Mehl, M. R., & Neiderhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54, 547-577.

[20] Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007*). Linguistic inquiry and word count: LIWC 2007*. Austin, TX: LIWC (www.liwc.net).

by the category: Thus "insult*" would also include the words "insulting" and "insults." For a given set of text, LIWC returns the percentage of words in the set that are contained in each of its word usage categories. In the case of Twitter messages, the content of a set of messages itself comprises the basic unit of text analysis. When the analysis, for example, returns a value of 2.3% words in the Anger category it means that of the total number of words in the document, 2.3% are words in the *Anger* word category.

We focused on a subset of the word categories contained in LIWC. These were word categories that psychological research on risk-related behavior suggests may be indicators of mental and emotional states conducive to risk-taking (Table 1). As part of the development process for LIWC, its authors obtained Base Rate values for word usage in each of the LIWC word categories. The Base Rates were obtained from multiple language generation sources, including Emotional Writing, Blogs, Novels and Talking. These values give an indication of percentage of word usages against which a specific context, such as Twitter content, can be contrasted. For our purposes, we utilized the Pennebaker, et. al., Grand Means for Base Rates comparisons in our results. In addition, we took advantage of a capability within LIWC to develop and apply a user-defined word usage category to account for alternative theoretical representations of linguistic expressions.

Table 1. LIWC word categories, categories example, total words in category and base rates.

Word Category	Examples	Total Words In Category	Base Rate (Grand Mean %)
Affective Processes	Happy, cried, abandon	915	4.41%
Positive Emotion	Love, nice, sweet	406	2.74
Negative Emotion	Hurt, ugly, nasty	499	1.63
Anxiety	Worried, fearful	91	0.33
Anger	Hate, annoy, mad	184	0.47
Sadness	Ache, grief, heartbreak	101	0.37
Cognitive Processes	Cause, know, think	730	15.37

Methods of Data Collection and Organization

Data for the research was drawn from Twitter messages collected via the public Twitter API using key hashtags for social events of interest. The data was collected in real time, as events occurred which offered opportunities to capitalize on current social trends and conditions. On the other hand, it made it difficult to obtain complete data sets for some events, particularly those that initiated quite quickly.

Events of special interest were those involving social protest or violence, as well as those involving political protest or reaction. A second category included social events in response to emergency conditions. A third category included non-violent/non-protest/non-emergency social events to serve as reference conditions.

Social Events Studied

Ferguson (#Ferguson)
On Saturday, August 9, 2014, Michael Brown, a young black man, was shot and killed by a white police officer in Ferguson, Missouri. The event initiated a period of civil unrest, social protest and violence that disrupted the community, leading to confrontations with police authorities. Twitter data was collected for multiple periods beginning on August 13th through November 24th of 2014. This is the event for which we have the most data.

Baltimore Riots (#Baltimoreriots, #Baltimoreprotest, #Baltimoreprotests, #Baltimoreuprising)
On April 12, 2015, Freddie Gray, a young black man, was arrested by Baltimore police on suspicion of carrying a conceal weapon. While in custody, he fell into a coma and was taken to a local hospital. On April 19th, Mr. Gray died, leading to public protests in Baltimore that on April 25th became violent. Twitter data was collected for April 28th, 2015

Charleston (#CharlestonShooting)
On the evening of June 17, 2015, nine people were killed by a lone gunman during a prayer service at a church in Charleston, South Carolina. One of the victims was Clementa C. Pinckney, the senior pastor and a state senator. Police arrested a young man who later confessed that he committed the shooting in hopes of starting a race war. The event led to generally peaceful response. Twitter data was collected for June 18, 19 and 21, 2015.

Block the Boat for Gaza (#BlockTheBoat)
Block The Boat for Gaza was a series of social protests that took place in the San Francisco Bay Area of California in response to Israel's invasion of Gaza. The focal point of the protests was the docking of an Israeli ship at Oakland, California. Protests generally took the form of several marches in downtown San Francisco, the first of which took place on Sunday, July 20, 2014. Subsequent marches grew in size and were generally peaceful. March coordinators announced late in July a protest gathering scheduled for August 16th, the date the Israeli ship was anticipated to dock. The resulting gathering confronted authorities, but not violently.

People's Climate March (#peoplesclimate)
The People's Climate March was a large-scale event to promote awareness and action regarding climate change. The march took place on Sunday, September 21, 2014 in New York City with companion marches worldwide. The estimated number

of participants was on the order of 300 – 400 thousand in New York City, making it one of the largest protest marches in history. The march was highly organized and focused on advocating change through peaceful protest.

Wildfires (#36PitFire, #KingFire)
Two wildfires, the "36 Pit Fire" and the "King Fire" occurred in Oregon and California respectively in 2014. The 36 Pit Fire started on September 13, 2014, east of Portland, Oregon, and burned over 5,000 acres. The King Fire ignited on the same day on the western slopes of the lower Sierra Nevada Mountains in California, ultimately burning over 97,000 acres and threatening several mountain communities. Both fires were human caused. Because both fires occurred around or near relatively populated areas, there was a significant social response.

Thanksgiving (#Thanksgiving) & Black Friday (#BlackFriday)
For reference purposes we collected data for two events that have a large social component but that do not involve social protest. These were Thanksgiving and Black Friday. For Thanksgiving (#Thanksgiving) we collected data for the day before Thanksgiving (November 26, 2014), as well as the day after. Black Friday (#BlackFriday) is has become a traditional holiday shopping event with a large social component. We collected data for the day before Black Friday (November 27, 2014), as well as two days after.

Data Organization

For each of the aforementioned events, Twitter data was collected in batches of 100,000 (100K) messages for #Ferguson, #Charleston, #Baltimoreriots, #Thanksgiving and #BlackFriday. For the remaining events, less data was available and the number of message per unit was smaller. Table 2 shows for each hashtag the number of 100K datasets as well as the number of tweets for those datasets having fewer than 100K tweets.

Table 2. Number of datasets for each hashtag studied.

Hashtag	Number of 100K Datasets	Number of <100K Datasets
#Ferguson	24	
#Charleston	3	
#BaltimoreRiots	1	
#BlockTheBoat	1	1
#PeoplesClimate	1	1
#Thanksgiving	4	
#BlackFriday	4	
#Wildfires		5

The data for #Ferguson was collected in two periods. The first period (early) began on August 14th, and ended August 22nd, 2014. The second period (late) began November 24th and ended November 27th, 2014. However for the first (early)

period the data was not continuous. For the early period, data was collected August 14th, 15th, 17th, 19th, 20th & 22nd, for a total of 18 100K datasets. For the late (second) period, data was collected on November 24th, 25th, 26th and 27th, for a total of 6 100K datasets. For all other hashtags (events) data was collected in continuous periods.

The basic unit of linguistic analysis was the 100K datasets. For these *100K sets*, only the content of the Twitter messages was used with all other information removed. Retweets were also removed.

For some analyses, the 100K datasets were broken down into *10K subsets*, thereby yielding a more precise picture of variation in network characteristics over time. The *10K subsets* were used in correlation analyses.

RESULTS

Affect, Sentiment and Risk Taking

Research on human judgment and decision making has revealed that the psychological processes invoked in uncertain situations rely heavily on affective experience as a mechanism for resolving complexity.[21] Indeed, one of the most fundamental psychological processes that people use to comprehend their world is *affective evaluation*.[22] Affect can be viewed as a quality assigned to an object, such as a plan of action, an opportunity or the outcome of a planned risk. In the context of risk, affective evaluation has been shown empirically to be a key element in risk perception.[23] In a study of risk-based decision making, MacGregor and colleagues showed that linguistic ratings of risky decision options predicted choices among a set of uncertain alternatives, with high-risk alternatives more likely to be chosen if they received more positive linguistic evaluations.[24] Thus, the affective valence as expressed in linguistic terms tended to be associated with a stronger risk-taking propensity.

It is evident from this line of research that language is the carrier of affect, and that underlying affect can be accessed and analyzed through the codification of linguistic expressions. We believe that this finding is critical to identifying linguistic markers in social media databases that are predictive of risk-taking behavior.

As we have discussed previously sentiment is an important element for predicting responses to social and political events from Twitter content. The use of the concept of sentiment in that research has tended to be somewhat general and could be more likened to mood than the specific affective valence associated with goals or objectives of interest (e.g., political candidates, leaders, parties).[25] The work discussed here explores a focused use of affect as a basis for identifying goals and actions associated with risk-taking in a social political context.

Affect Diffusion

Affect diffusion is generally considered a form of social influence that involves the transfer of affective states from one individual to another. Thus, a person can acquire affective states such as anger, sadness, joy and anxiety from those they are connected to through social networks. This process has been defined more precisely

[21] Slovic, P., Finucane, M. L., Peters, E. & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, 24, 311-322.

[22] Damasio, A. R. (1994). *Descartes' error: Emotion, reason and the human brain.* New York, NY: Putnam & Sons.

[23] Slovic, P. (2010). *The feeling of risk: new perspectives on risk perception*. Washington, D.C.: Earthscan.

[24] MacGregor, D. G., Slovic, P., Dreman, D., & Berry, M. (2000). Imagery, affect, and financial judgment. *The Journal of Psychology and Financial Markets*, 1(2), 104-110.

[25] The exception to this is perhaps the work of Bollen, Mao, & Zeng (2011) discussed in Section 1.4.2. Their predictive framework for predicting stock market changes from Twitter discourse include a number of sentiment dimensions that are similar to affective scales.

by Peters & Kashima (2015) as ". . . a process whereby one person's affective action—that is, an action that reflects or provides information about his or her current affective state—leads another person to experience a congruent affective state" (pg. 968).[26] Others have focused on the notion of *collective emotions* that result from " . . . ". . . the synchronous convergence in affective responding across individuals towards a specific event or object" (p. 406).[27]

While affect diffusion is fairly well established in laboratory contexts, its presence in social media and social media networks is less clear. However, in an empirical study of "emotional contagion" in the context of Facebook, researchers found clear evidence that exposure to experimentally-induced emotional expressions influenced people to post content that was consistent with the affective valence of the exposure.[28]

Anger and Risk

Although anger is a negative emotion, it has some unique characteristics with respect to its influence on how people appraise and predict the future, including the outcomes of their actions in risky situations. In general, negative emotions tend to instill a pessimistic view that leads to conservative expectations with regard to outcomes of risk taking. However, anger tends, despite its negativity, to bias expectations toward positive outcomes and a tendency toward risk-seeking.[29] Thus, we see in the context of judgment and choice under uncertainty a tendency toward emotion specificity, with anger operating in the opposing direction of its negative valence.[30] Other research has examined anger and sadness with respect to their differences in terms of action tendencies and related behavioral implications, with sadness having a tendency to attenuate actions and anger having the opposite effect.[31] The results imply that anger and sadness have contrasting effects with respect to behavior, with anger offering opportunities for mastery and control and sadness inhibiting actions and promoting withdrawal.

[26] Peters, K. & Kashima, Y. (2015). A multimodal theory of affect diffusion. *Psychological Bulletin*,141, 966–992.
[27] von Scheve, C., & Ismer, S. (2013). Towards a theory of collective emotions. *Emotion Review*, 5, 406-413.
[28]. Kramer, A .D. I., Guillory, J. E., & Jeffrey T. Hancock, J. F. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences (PNAS),* 111, 8788-8790.
[29] Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, 81, 146-159.
[30] Lerner, J. S., & Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognition and Emotion*, 14, 473-493.
[31] Mouilso, E., Glenberg, A. M., Havas, D. A., & Lindeman, L. M. (2007). Differences in action tendencies distinguish anger and sadness after comprehension of emotional sentences. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1325-1330). Austin, TX: Cognitive Science Society.

Linguistic Expression of Affect

We begin by examining message content with respect to word categories that reflect affect and emotion, for the various events studies as well as the average for all of the Ferguson datasets (*100K sets*).[32] For Thanksgiving and Black Friday, the percentages of Affect words generally met or exceeded the base rate while for Charleston and Ferguson rates were near or slightly higher than the base rate, but not as high as Thanksgiving.

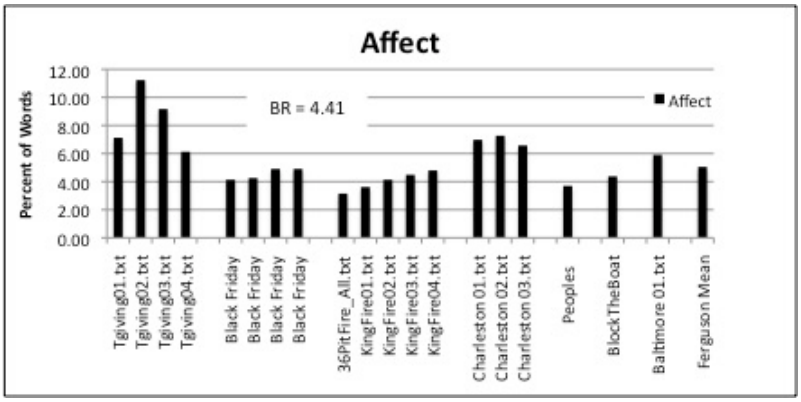

Figure 1. Distributions of Affect word usage by event.

If we look separately at the positive and negative emotion words that comprise the larger Affect Words category, we see that for Thanksgiving and Black Friday, positive emotion words dominated negative emotion words by a sizable proportion, and well above the base rate for positive emotion words (Figure 2). For all of the other events, the difference between percentages of positive and emotion words were much small and even negative, particularly for the disruptive social events Baltimore and Ferguson as well as Charleston.

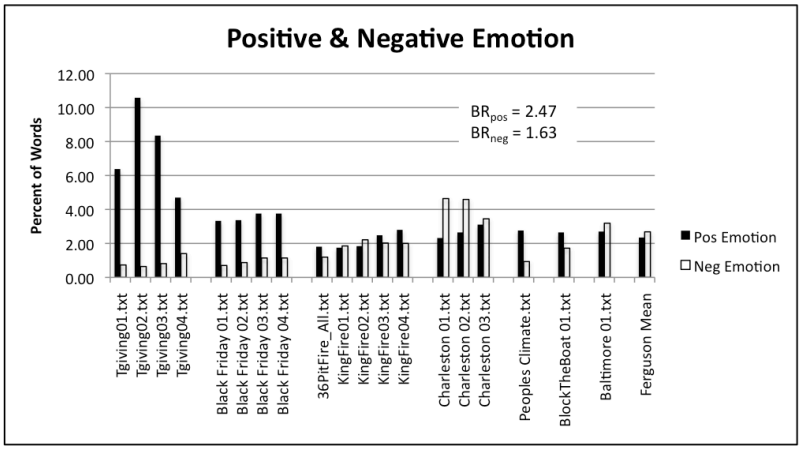

Figure 2. Distributions of Positive and Negative Emotion word usage by event.

In general, word generation for the various base rate categories reported by Pennebaker, et. al. are biased in a positive direction. That is, people tend, across a variety of word use situations, to use more positive words than negative words by about +0.84%. We can take the different between positive and negative word usage percentages as a measure of sentiment or relative affect (Figure 3). Charleston,

[32] All analyses in this section are based on *100K sets* of Twitter messages.

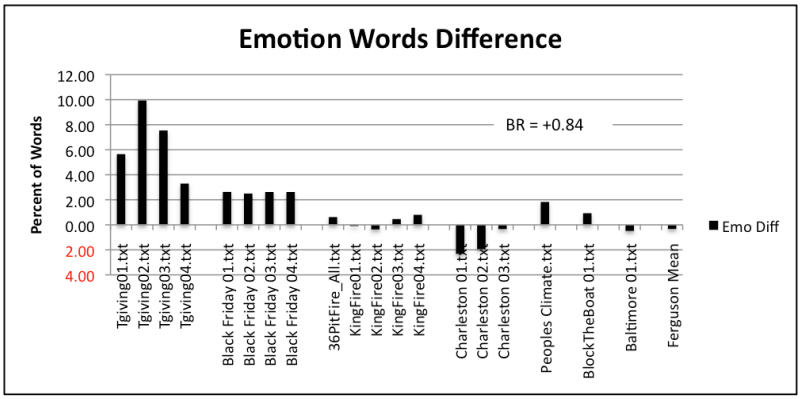

Figure 3. Distribution of Positive and Negative Emotion word usage difference by event.

Baltimore and Ferguson all exhibited (on average) negative sentiment while the social events Thanksgiving and Black Friday trended quite positive sentiment.

Turning specifically to Ferguson, Affect word usage was fairly consistently at or above the base rate across all of the datasets obtained. Some of the highest periods were those associated with Ferguson after the Grand Jury announcement (Figure 4). Use of negative emotion words was above the base rate for all of the datasets, while positive emotion words exceeded the base rate for only three of the datasets (Figure 5).

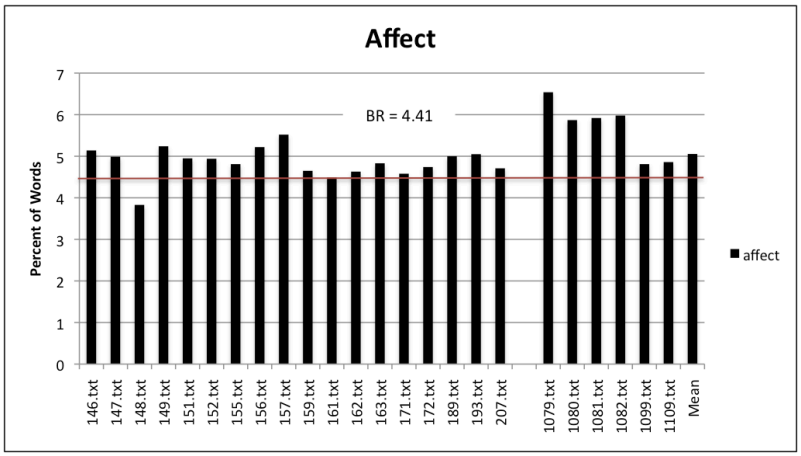

Figure 4. Distribution of Affect word usage for Ferguson event.

In terms of the difference between percentages of positive and negative words, for all but four of the datasets the difference was negative. Of the four instances for which the difference was positive, for none of the four did it reach the base rate (Figure 6).

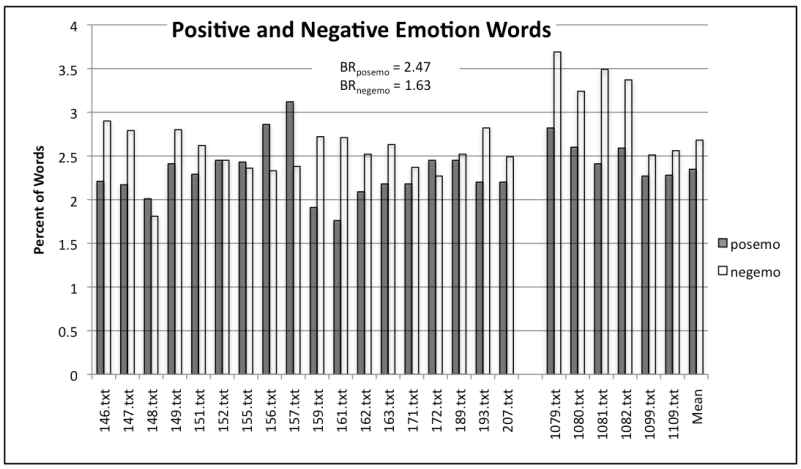

Figure 5. Distributions of Positive and Negative Emotion word usage for Ferguson event.

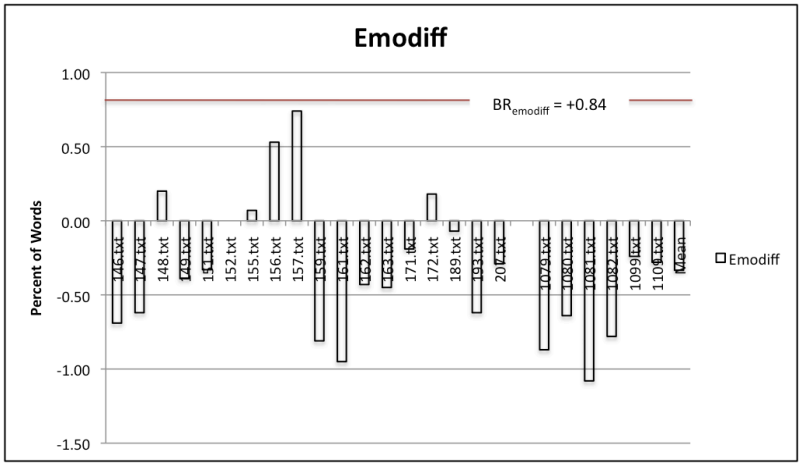

Figure 6. Distribution of Positive and Negative Emotion word usage difference for Ferguson

Overall, Affect word usage in Ferguson Twitter content for the datasets studied did not exhibit any particular remarkableness overall. However, when separated into positive and negative emotion words, a different picture emerges. Sentiment, as measured by the difference between percentages of positive and negative emotion words, was generally quite negative with few positive excursions, and then not sufficient to reach base rate.

Anger, Anxiety and Sadness Word Usage.

Previously we discussed the relationship of anger to risk. Essentially, anger can heighten risk taking for several potential reasons. First, as an emotion it can be highly energizing and may create conditions conducive to taking actions. Second, it can be very goal-directed in the sense that it may be based on resentment toward specific social objects such as authority figures. Third, expressing and/or acting on anger can give rise to feelings of control and mastery. In general, as reflected in Pennebaker et. al. base rates, Anger word usage is relatively low across a range of contexts at 0.47%. Likewise, sadness also tends to have a low base rate at 0.37%, for a ratio of 1.27.

Looking across all social events studied, we see some sharp disparities, particularly with respect to Anger word usage but also Sadness word usage though to a slightly lesser degree (Figure 7). Anger word usage was particularly high for the Charleston event as well as for Baltimore and Ferguson Mean. Anxiety word usage tended to be highest for the Charleston event.

These trends are perhaps reflected better in the Anger/Sad Ratio where departures from base rate (BR = 1.27) are very high, particularly for Charleston, Baltimore and Ferguson Mean (Figure 8). The events Thanksgiving, Black Friday and even the Wildfire event are generally close to base rate for Anger/Sad Ratio.

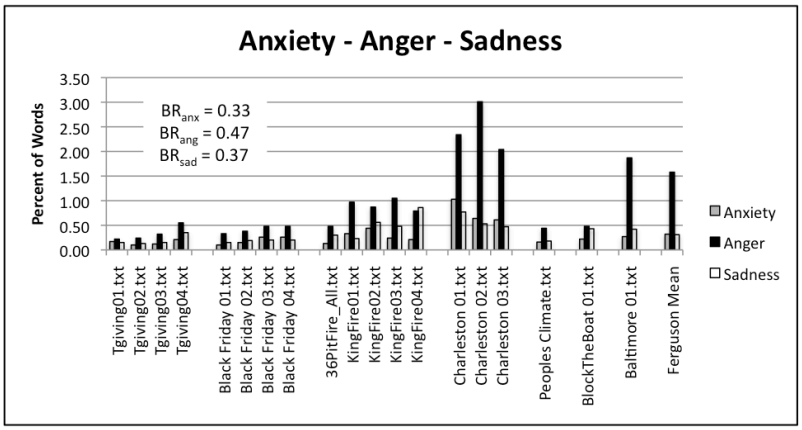

Figure 7. Distribution of Anxiety, Anger and Sadness word usage by event.

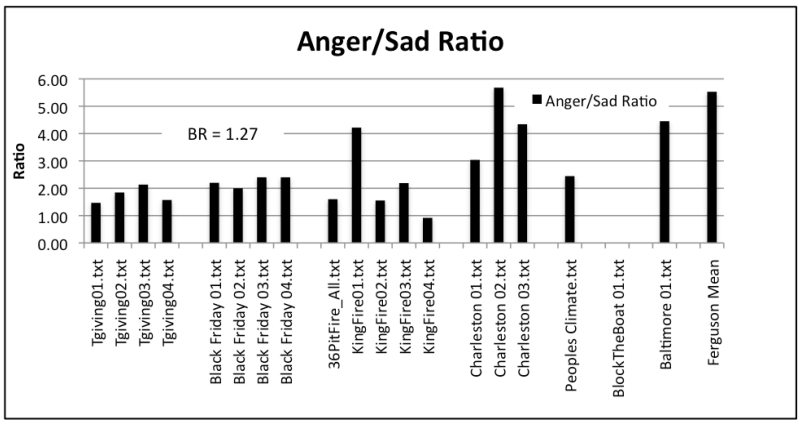

Figure 8. Distribution of Anger/Sad Ratio by event.

Turning to Ferguson in greater detail, we see that Anger word usage was generally high and well above base rate for all of the datasets collected, and particularly after the Grand Jury Decision where some of the highest Anger word usage levels were observed (Figure 9).

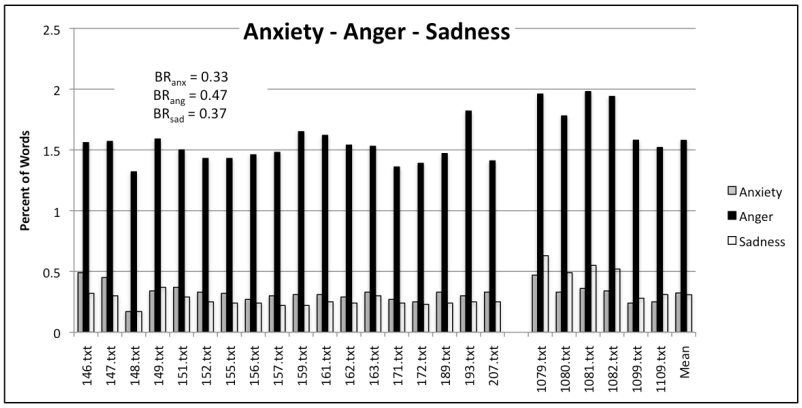

Figure 9. Distribution of Anxiety, Anger and Sadness word usage for the Ferguson event.

Also observed were some of the highest levels of Sadness word usage, particularly after the Grand Jury Decision. The Anger/Sad Ratio shows that for most of the Ferguson event prior to November, 2014, the Ratio was quite high when compared with base rate, and with considerable fluctuation (Figure 10). After the Grand Jury Decision, the Anger/Sad Ratio was at its lowest level but trended upward through the remaining datasets.

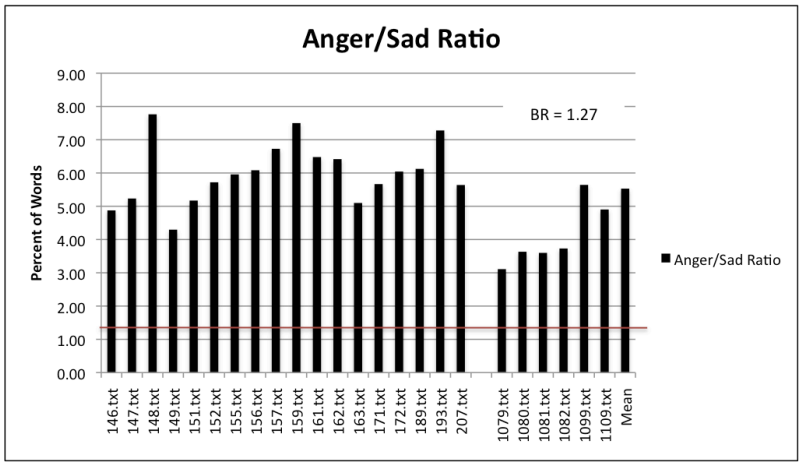

Figure 10. Distribution of Anger/Sad Radio for the Ferguson event.

Cognitive Mechanisms and Affect.

The word category *Cognitive Mechanisms* is relevant because it captures the tendency toward word usage that reflects thought and reasoning. The general base rate for Cognitive Mechanisms word usage reported by Pennebaker, et. al, is 15.37%. In our analysis, we contrast Cognitive Mechanisms word usage with Affect word usage as a relative measure of the degree to which message content is indicative of reasoning processes versus affective or emotional processes.

Across all of the datasets studied Cognitive Mechanisms word usage was below the base rate (Figure 11). Highest levels of Cognitive Mechanisms word usage were for the Charleston event, as well as Baltimore and Ferguson.

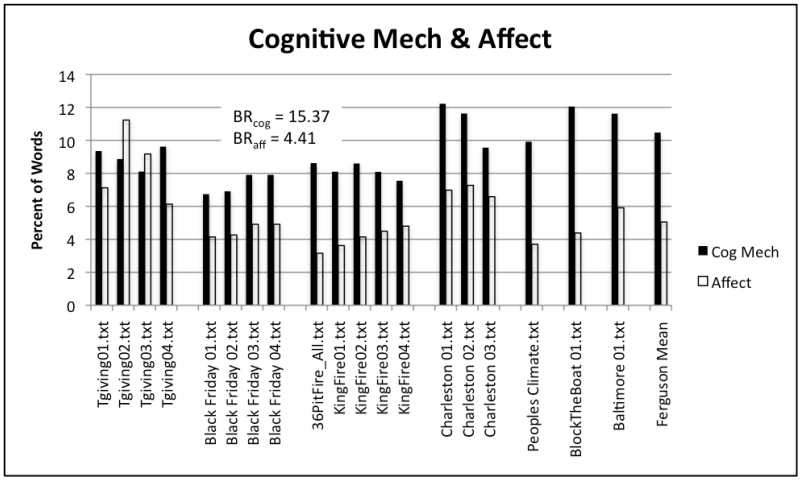

Figure 11. Distributions of Cognitive Mechanisms and Affect word usage for all events.

The ratio of Cognitive Mechanism to Affect word usage provides a relative indication of the relative balance of cognitive to affective content (Figure 12). The base rate for the Cognitive/Affective Ratio is 3.49 based on Pennebaker et. al. base rates for each of the two word usage components respectively. The ratio was highest for the wildfire events as well as the social protests Block The Boat, Baltimore and Ferguson Mean. However, for none of the events did the ratio reach the base rate.

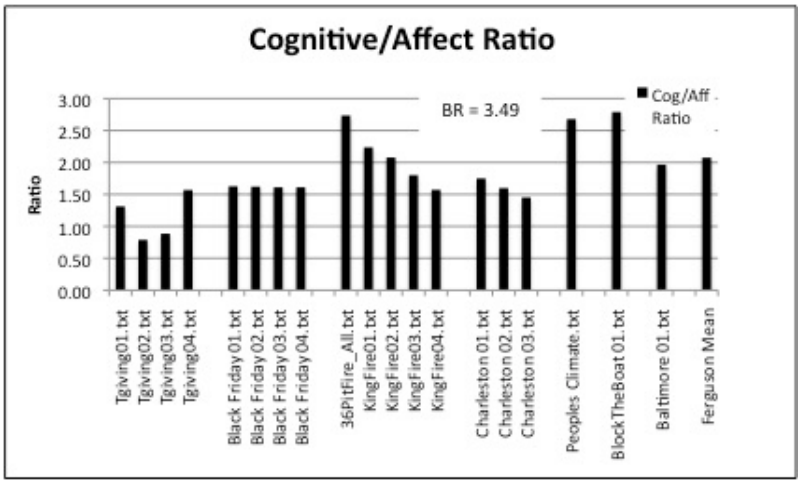

Figure 12. Distribution of Cognitive /Affect Ratio for all events.

For Ferguson (all datasets) Cognitive Mechanisms word usage was consistently below the base rate (Figure 13). High periods were exhibited at the time of the Grand Jury announcement, as well as early on in the event during times of disruption and violence.

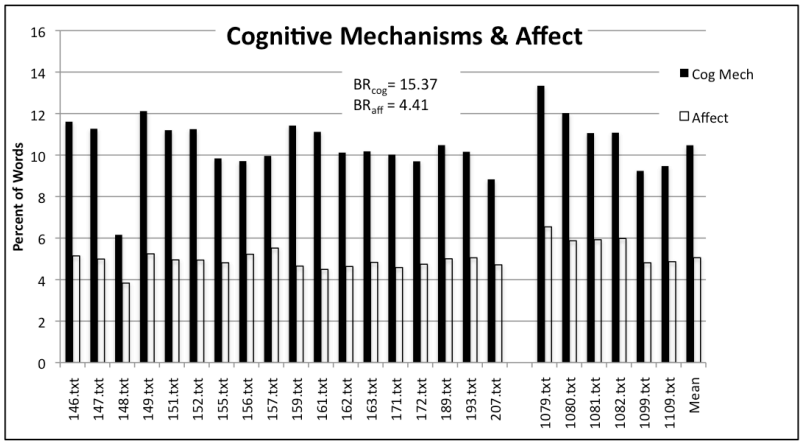

Figure 13. Distributions of Cognitive Mechanisms and Affect word usage for Ferguson.

The Cognitive/Affective Ratio was well below base rate for all datasets and exhibited some variability across the event (Figure 14).

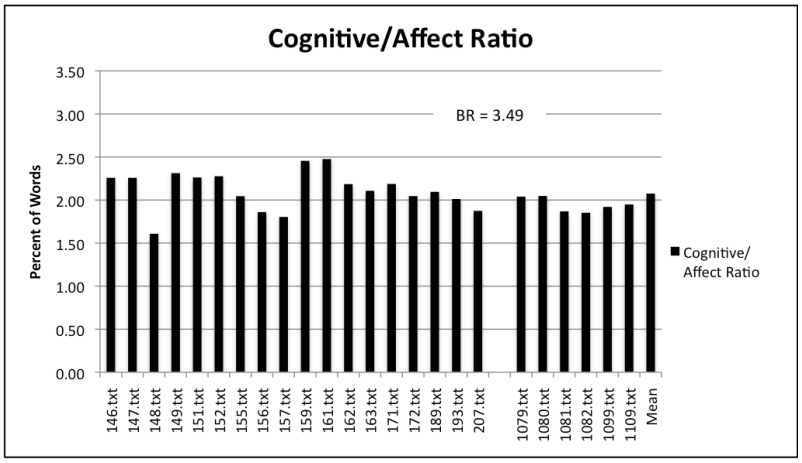

Figure 14. Distribution of Cognitive/Affect Ratio for Ferguson.

Cognitive Mechanism word usage tended to be relatively low compared to base rate for social events studied. We speculate that this tendency may be characteristic of Twitter content associated with the constrained length of Twitter messages (i.e., 140 characters). It may also reflect a tendency for cognition-related words to be longer than affect words, though we have not explored this feature of LIWC to determine if it is so.

Loss and Risk Taking: Redemptive Framing

Numerous studies in psychology have shown a profound asymmetry in people's responses to losses and gains.[33] In general, and across a very broad range of decisions involving uncertainty, people will prospectively avoid behavior that exposes them to the potential for loss if the opportunity is available.[34] Retrospectively, the experience of losses can be particularly distressing, and dominates the experience of gaining, which extends to include negative emotional reactions and mood shifts. [35]

As both a theoretical and practical matter, the experience of actual losses has relevance to risk-taking behavior. This can be seen in one of the most important theoretical frameworks to emerge from behavioral decision theory, that of *Prospect Theory*.[36] Prospect Theory builds upon a number of empirical findings relating to how people evaluate information, including that pertaining to risk and uncertainty. The theory is based on the asymmetry of gains and losses discussed above, in that losses are experienced more extremely than are gains for an equivalent objective change in one's position from a given reference point (Figure 15).

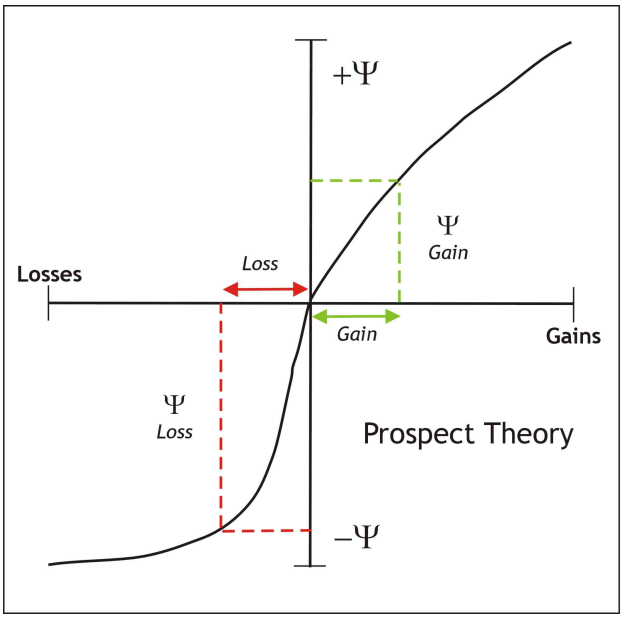

Figure 15. Graphic representation of *Prospect Theory*. Adapted from Kahneman & Tversky (1979).

Because the non-linear psychological loss function is steeper than the gain function, loss incurs a greater *emotional cost* per unit of change from a reference point or prior set of conditions. When the loss is experienced, the emotional cost is

[33] For a review see: Eldad, Y., & Hochman, G. (2013). Losses as modulators of attention: Review and analysis of the unique effects of losses over gains. *Psychological Bulletin*, 139(2), 497-518.
[34] Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323-370.
[35] McGraw, A. P., Larsen, J. T., Kahneman, D., & Schkate, D. (2010). Comparing gains and losses. *Psychological Science*, *21*, 1438-1445.
[36] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

borne, and the potential for further emotional loss is attenuated. The result is to provoke a risk-taking (or risk-seeking) tendency *to recover* to one's previous position. Prospect theory has been found to account for a wide range of differential responses to situations that involve loss compared with those that involve gain, including the sunk cost effect whereby people will incur much greater potential for loss to recover from a failed condition than they would to initiate the condition having no prior investment.[37,38]

The implications of Prospect Theory for understanding group and individual dynamics involving risk-taking behavior reside in the importance of the perception of loss as a motivator for risk-taking behavior, as well as the concept that part of the experience of loss is a desire to return to a pre-existing set of conditions that represent a reference point of particular value. So, for example, in the case of groups or individuals who hold conservative political or religious views, changes brought about by more modern conditions may be perceived as a loss relative to the reference point of an earlier and more desirable time. Framed this way, risk-related actions or behaviors may shift toward risk-taking in the interests of recovering that loss. The same line of reasoning can be applied to other social or political circumstances where the current state of a group or an individual is seen as a loss when compared with some other idealized state or condition, resulting in risk-taking behavior to bring about a recovery to that state.

We explored linguistic expressions of loss recovery that have an association to risk-taking behavior along the lines predicted by Prospect Theory. Our line of examination was based on the notion of *redemptive framing*. Redemptive framing refers to linguistic markers that are indicative of a motivation to redeem current conditions to a more desirable set of ideal conditions. Linguistic markers consistent with redemptive framing include: *recover, regain, get back, retrieve, undo, return, save and rescue*. Our expectation was that the presence of language consistent with redemptive framing in Twitter message content will provide an indicator of the presence of conditions conducive to risk-taking behavior, particularly in social contexts involving social and political reaction to change.

Linguistic Expression of Redemptive Framing

Redemption is essentially a moral claim to something seen as rightfully belonging to one but which has been lost through some act of injustice and/or injury. Its expression linguistically can take various forms that relate to a common theme associated with getting back something that has been let go, lost or taken. Expressions based on this theme may appear in association with strongly affective words or expressions signifying anger, frustration or righteous indignation.

Our approach was to develop a word usage concept based on redemptive framing by identifying words and word passages that have strong association to the concept of redemption. Those words and word phrases having the strongest

[37] Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York, NY: Cambridge University Press.

[38] Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York, NY: Cambridge University Press.

association were used as a dictionary element to apply to Twitter content, using the LIWC software capability to apply an external dictionary.[39]

Redemption word usage was computed for all of the datasets as well as Ferguson Mean. For holiday-related social events, Redemption word usage was generally low, and in the range of approximately 0.20% for Thanksgiving and up to 0.5% for Black Friday (Figure 16).

Unlike the word usage concepts that are part of the LIWC package, *Redemption* does not have a base rate such that the rates observed can be compared with a broad standard. To some degree the rates observed for Thanksgiving and Black Friday may represent a lower range for social events as expressed in Twitter content.

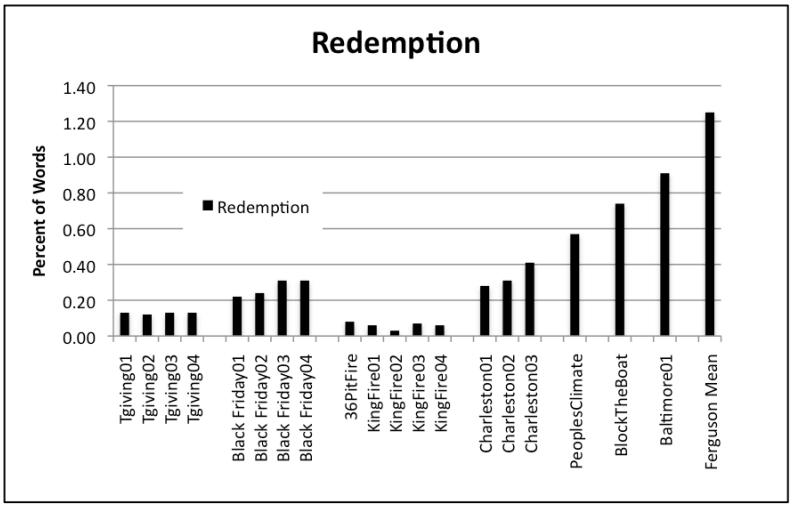

Figure 16. Distribution of Redemption word usage by event.

For the various Wildfire events, Redemption word usage was the lowest of any event. On the other hand, for the disruptive social events Redemption word usage was much higher, particularly for Peoples Climate, Block The Boat, Baltimore and Ferguson Mean.

Looking at Ferguson in greater detail, Redemption word usage varied above and below the average for the entire event for which datasets were analyzed. After the Grand Jury Decision, Redemption word usage remained at the average or above through the last dataset collected (Figure 17).

[39] The method of creating the *Redemption* word usage category followed along the lines of that reported by Pennebaker, et. al., for their creation of the word usage categories in LIWC.

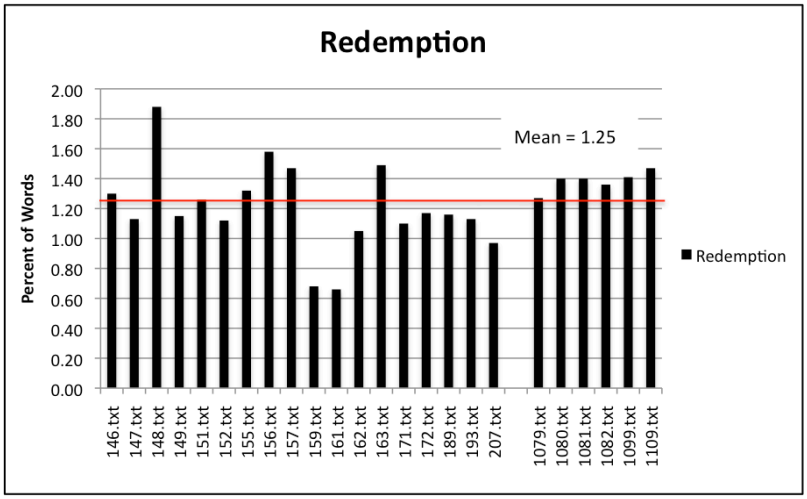

Figure 17. Distribution of Redemption word usage for the Ferguson event.

We also examined some internal relationships between Redemption word usage and other key word usage categories as well as word usage category ratios. One could argue that Redemption would bear a moderate to strong relationship to Anger. However, with a correlation between Anger word usage and Redemption of only r = +0.103, this speculation is not borne out in the data for Ferguson (Figure 18).[40]

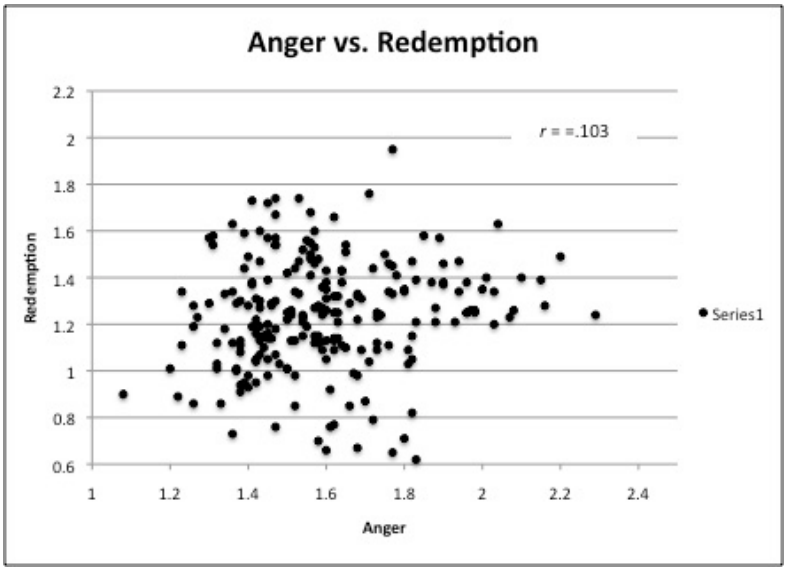

Figure 18. Scatterplot of Anger and Redemption word usage for the Ferguson event (r = 0.103, df = 229).

Alternatively, one could hypotheisze that Redemption would perhaps bear at least a moderate relationship with Affect word usage, arguing that emotion would activate language usage consistent with calling for a return of what has been lost or foregone. Indeed, a moderate correlation between Affect word usage and Redemption was present at r=0.375, indicating greater Affect word usage associated with higher levels of Redemption word usage (Figure 19). Likewise, one might

[40] Scatterplots are based on *10K subsets* of the *100K sets*.

speculate an inverse correlation between Cognitive Mechanism word usage and Redemption based on the hypothesis that higher levels of reasoning and logic would translate into less provocative language use. Upon examination, however, the direction of the relationship appears correct, but the strength appears quite low at r=-0.122.

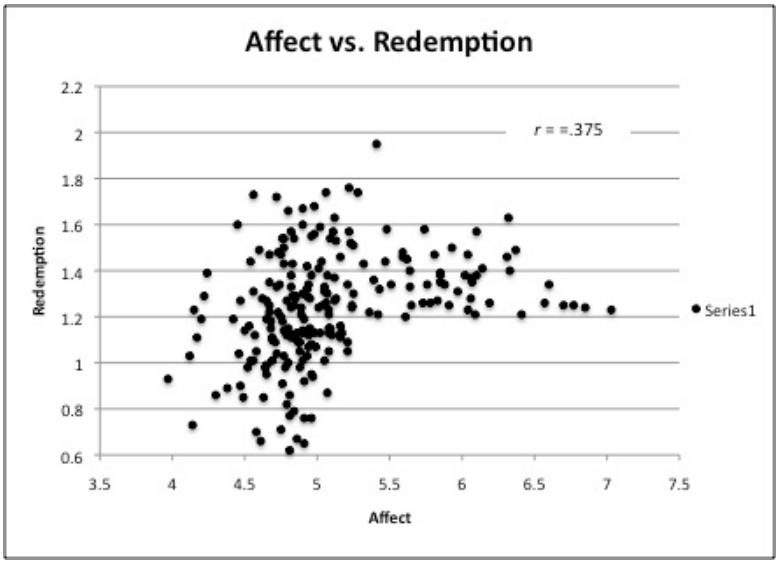

Figure 19. Scatterplot of Affect and Redemption word usage for the Ferguson event (r = 0.375, df = 229).

A third hypothesis is that the relationship between Redemption and Cognitive Mechanisms and Affect is more a matter of the balance between "feeling" and "thinking." We captured this balance in terms of the Cognitive/Affect Ratio which does bear a strong inverse relationship with Redemption at r=-0.547 (Figure 20).

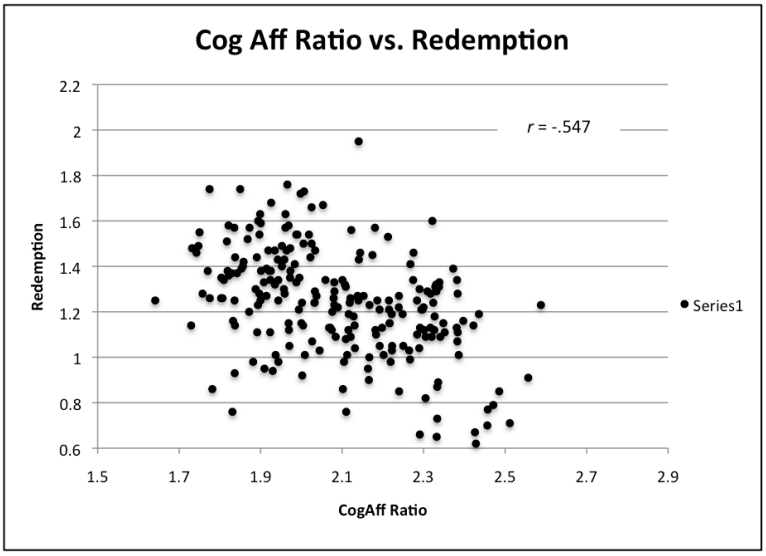

Figure 20. Scatterplot of Cog/Aff Ratio and Redemption word usage for the Ferguson event (r = -0.547, df = 229).

Linguistic Markers, Network Centrality and Risky Shift

For a number of decades, behavioral and social science researchers have had an interest in a phenomenon known as *risky shift*. Although *risky-shift* was the phenomenon that early-on led to the identification of the broader issue of group polarization, much of the research done on group polarization has focused on changes in attitudes and opinions as a function of group formation and topic discussion. While attitudes are an important precursor to behaviors, we highlight the importance of risk-taking behavior as a focal issue of interest. Risky-shift refers to a group-induced tendency toward risk taking, and the general observation over years of study is that individuals have a greater tendency to takes risks in a group context than they do as individuals.[41,42,43]

Social dynamics are a critical element in predicting risk-taking behavior based on the notion of risky shift. Twitter has been studied in terms of its properties as an indicator of social dynamics, and particularly the information that can be gleaned from the analysis of tweets concerning the phenomenon of *group polarization*. Group polarization refers to a tendency for groups to hold more extreme positions or attitudes than the average of the pre-group attitudes of its individual members. In cases where the attitudes of individuals already tend toward cautiousness, group polarization can tend to amplify that caution. On the other hand, when attitudes tend toward less cautious or risky, polarization can lead a group toward greater risk. The phenomenon of group polarization is important because it helps explain human behavior in a of real-life situations, such as policy decisions.[44] Moreover, these effects have been observed and demonstrated in cultures around the world with varying types of group participants.[45]

Key to group polarization is the interaction of its members in some form of discourse.[46] In the context of computer-mediated communication Sia, Tan & Wei found that even in contexts where only text interaction occurred (i.e., no visual cues) group polarization was not only present, but even more extreme than in cases where the bandwith of interaction was greater (i.e., visual cues).[47] Some commentators have noted that the combination of modern media and the Internet have exacerbated a trend toward polarization that has become more extreme simply

[41] Myers, D. G. & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83, 602-627.

[42] Begum, H. A. & Ahmed, E. (1986). Individual risk taking and risky shift as a function of cooperation-competition proneness of subjects. *Psychological Studies*, 31, 21-25.

[43] Crott, H. W., Szilvas, K., & Zuber, J. A. (1991). Group decision, choice shift, and polarization in consulting, political, and local political scenarios: An experimental investigation and theoretical analysis. *Organizational Behavior and Human Decision Processes*, 49(1), 22-41.

[44] e.g., Whyte, G., & Levi, A. S. (1994). The origins and function of the reference point in risky group decision making. The case of the Cuban missile crisis. *Journal of Behavioral Decision Making*, 7, 243-260.

[45] e.g., Forsyth, D. R. (1990). *Group dynamics*. Pacific Grove, CA: Brooks Cole Publishing.

[46] Van Swol, L. M. (2009). Extreme members and group polarization. *Social Influence*, 4(3), 185-199.

[47] Sia, C. L., Tan, B., & Wei, K. K. (2002). Group polarization and computer-mediated communication: Effects of communication cues, social presence and anonymity. *Information Systems Research*, 13(1), 70-90.

because individuals can seek out the attitudes of others with similar views, thereby increasing their confidence in their own opinions.[48]

Taken together, these research findings indicate that group polarization can occur without physical engagement, and can be even more potent as a basis for group formation in situations where individuals can self-select themselves anonymously into groups based upon their own prior attitudes and positions. Indeed, the picture that emerges with respect to microblogs is that Twitter can readily foster and sustain the growth of group polarization through the consistent dialogue that is part of "tweeting", no matter what their geographic location.[49].

Graph Generation

To examine the social structure of Twitter messages, we used methods based on graph generation. The purpose of graph generation is to extract and map the embedded relationships within Twitter messages modeled as a *social interaction graph*. The graph is used to identify and compute metrics that are then applied to investigating network structure. We applied five graph metrics to social interaction graphs.[50] These are:

- Degree Centrality: A network is comprised of nodes and their linkages or "edges." Nodes that are more central in a network than other nodes have higher *centrality*, which is measured by how many other nodes are adjacent or connected to it.
- PageRank Centrality: PageRank centrality is similar to Degree centrality but considers both the number of links to any node as well as the importance of each of the nodes connected to it. In this regard, PageRank centrality roughly represents how central (or important) a given node is to the overall graph.
- Average (Global) Clustering Coefficient: Networks differ in the degree to which their nodes cluster together. Networks in which two connected nodes are also connected to the same third node exhibit more clustering than a network in which two connected nodes are not connected to the same third node.
- Transitivity: The transitivity metric is closely related to the Average Clustering Coefficient but only considers the triads – that is, two edges with a shared node. Both the Average Clustering Coefficient and Transitivity describe the degree to which a network exhibits clustering. Higher values of transitivity indicate a greater tendency for the members of a network to be connected.

[48] Sunstein, C. (2008). The law of group polarization. In J. S. Fishkin & P. Laslett (Eds.), *Debating deliberative democracy*. Blackwell Publishing, Ltd.

[49] Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, 30(5) 316-327.

[50] See Appendix A for a description of the mathematics and related formal definitions for the various graph metrics discussed.

- Giant Component: A giant component refers to the percentage of nodes in a network that are connected to the largest subgraph (set of connected nodes) from the original network. As the network is more connected, the giant component score increases.

Relationship of Linguistic Categories with Percentage of Giant Component

In this section we present an overview of relationships between key linguistic categories and the network measure Percentage of Giant Component. In general, a component is a subgraph of a network in which all nodes are reachable from all other nodes in the subgraph[51]. A given network may have one or more subgraphs or subcomponents. As we defined previously, the Percent of Giant Component is the percentage of nodes connected to the largest subcomponent of a network graph. An example is illustrated in Figure 1 in which all of the nodes shown are connected to a single subgraph.

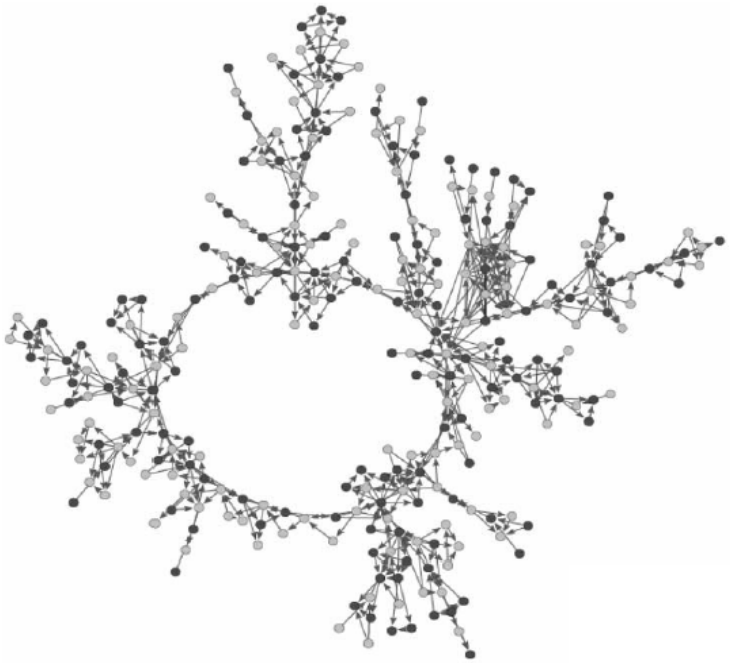

Figure 21. Example of a Giant Component in which every node is reachable by some path from every other node. (Adapted from Bearman, Moody & Stovel, 2004).[52]

The choice of a network metric should be based, at least in part, on what one either believes or speculates is transmitted or propagated through the network. In

[51] Wasserman, S., & Faust, K. (1994). *Social Network Analysis*. Cambridge: Cambridge University Press.
[50] Bearman, Moody & Stovel (2004).

our analyses of linguistic content of Twitter messages, the conceptual framework within which our research has taken place is that of affect and emotion as a core driver of risk-taking propensities. Given our previous discussion of affect as a "diffusion" process through social networks, the choice of the Giant Component is due to its close relationship to the application of component analysis in other domains, such as infectious disease, where diffusion (i.e., infection) is of primary interest.

We focus on the Giant Component metric and its relation to the Ang/Sad Ratio discussed previously. The Ang/Sad Ratio is the ratio of the percentage of *Anger* word usage to the percentage of *Sadness* word usage (Figure 22).

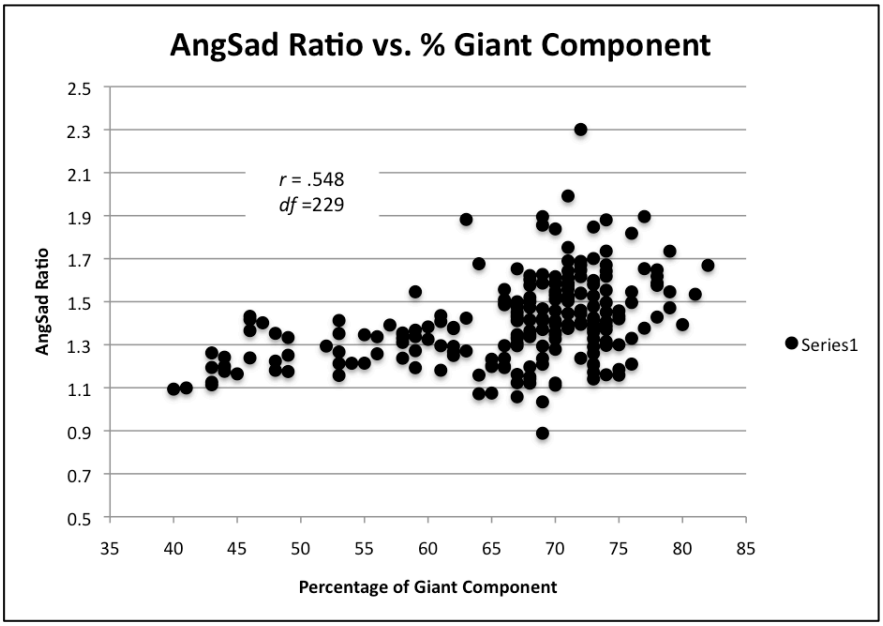

Figure 22. Scatterplot of Ang/Sad Ratio vs. % Giant Component for Ferguson event.

The observed correlation is quite high ($p<.001$) and shows a strong positive association between Ang/Sad Ratio and % Giant Component. During periods when the Ang/Sad was relatively large, the % Giant Component also tended to be larger. The result suggests that there is an association between the linguistic content of Twitter messages in terms of the ratio of Anger word usage to Sadness word usage and the connectedness exhibit in the network as measured by the percentage of nodes connected to the largest sub-component of the network.

We see a similar result when we examine the relationship of the Cog/Aff Ratio with the % Giant Component. The Cog/Aff Ratio is the ratio of the percentage of Cognitive Processes word usage to the percentage of Affective Processes word usage. Essentially, the Cog/Aff ratio is the relative balance of cognition to emotion. Whereas the Ang/Sad Ratio is the relative balance of two distinct categories of psychologically offsetting emotions. Here we see a slightly more complex picture (Figure 23).

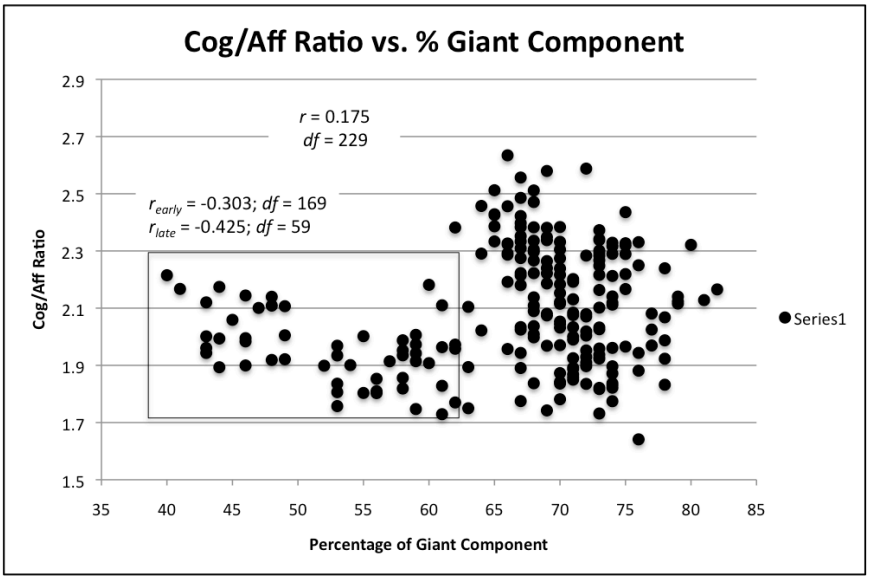

Figure 23. Scatterplot of Cog/Aff Ratio vs. % Giant Component for Ferguson event.

Looking at the whole scatterplot, it appears that there is a slight positive correlation (r = .175; p<.01) between Cog/Aff Ratio and % Giant Component. However, when the datapoints are separated into those due to the early phase of Ferguson (Aug 14 – 20, 2014) and the late phase of Ferguson (Nov 23 – 24, 2014) after the Grand Jury Decision a different picture emerges. Datapoints for the late phase of Ferguson are shown in the box imposed on the scatterplot. Correlations between Cog/Aff Ratio and % Giant Component are, under these conditions, inverse for both the early phase (r = -.303; p < .0001) and the late phase (r = -.425; p < .001). This indication suggests that as the relative balance of cognition and emotion tips more in the direction of emotion, as reflected in a lower Cog/Aff Ratio, there is an associated increase in the size of the Giant Component.

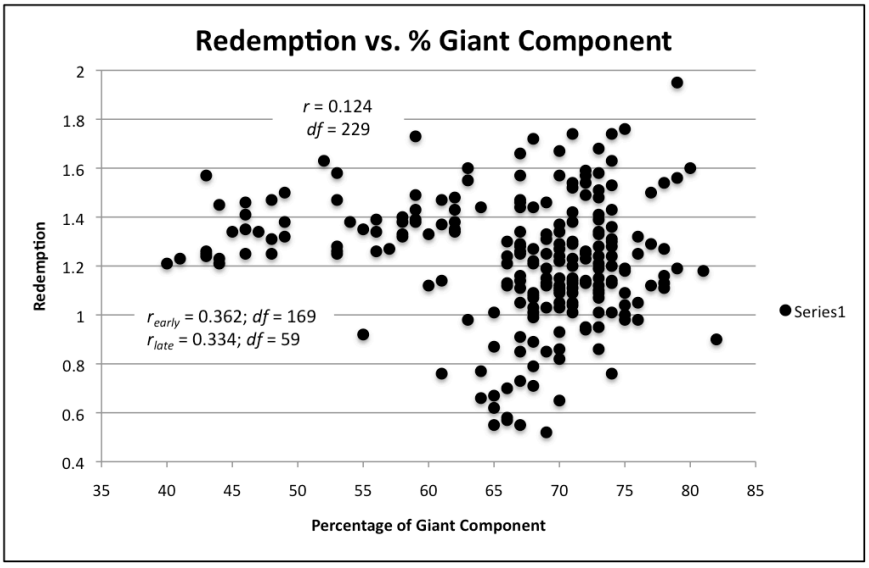

Figure 24. Scatterplot of Redemption vs. % Giant Component for Ferguson event.

Lastly, we examine the relationship between Redemption (Redemptive Framing) and the % Giant Component (Figure 24). The correlation between Redemption and % Giant Component, computed using all of the datapoints, was a modest $r = .124$ ($p < .05$). However, splitting the data into Ferguson early and Ferguson late resulted in larger correlation coefficients for both subsets with Ferguson early ($r = .362$; $p < .001$) and Ferguson late ($r = .334$; $p < .001$) nearly equal in magnitude and positive in direction. The result suggests a positive association between Redemption and % Giant Component: Greater Redemption word usage bears a correspondence with a larger % Giant Component.

Overall, the results suggest a picture of connectedness within the Ferguson Twitter network that exhibits a relationship with language usage in terms of word categories consistent with psychological concepts. These psychological concepts include both cognition and emotion, and themselves have relationships with risk-related behavior. Our analyses of these concepts as they relate to connectedness suggest that conditions conducive to risk-taking may be detectable by linguistic analysis of both Twitter content as well as network metrics that could provide sufficient group identity to provoke group-related risk behavior.

Key Findings

The basic question that motivated this research hinged on the potential of methods for analyzing social media, based on the psychology of risk, to identify conditions conducive to risk-taking behavior by examining a set of social events that range in social disruptiveness. To address this question, we applied two basic methods of analysis: one that targets language use in Twitter message and that conceptualizes language as the expression of cognitive and emotional conditions, and a second that uses network metrics to identify relative presence of connected conditions between network entities reflect relative degrees of social organization. To these ends, we have found the following:

1. Across a range of social events, some of which exhibited extreme social behavior with respect to disruption and violence and others of which were more benign, language usage varied along lines that reflect the presence of conditions conductive to greater degrees of risk-taking propensity. Key word usage categories using the LIWC program for text analysis revealed high levels of word usage for the word categories Affect and Anger, as well the ratio of *Anger* to *Sadness* that generally corresponded to increasing levels of social disruption.

2. Predictions of risk-related behavior based on Prospect Theory and codified in terms of a word usage concept labeled *Redemption* was found to distinguish between social events having high and low social disruptiveness. Social events with high disruptiveness (e.g., Ferguson, Baltimore) evidenced higher levels of *Redemption*, whereas social events low in disruptiveness (e.g., Thanksgiving) showed lower levels of *Redemption*.

3. *Redemption* also related strongly to some word usage categories from the LIWC category set, but not others. For example, the relationship between *Redemption* and *Anger* was marginal, but relationship with *Affect* was high and particularly high with the ratio of *Cognition* to *Affect* (emotion). So, for example, looking within the dataset for the Ferguson event, the correlation was high and inverse between the ratio of *Cognition* to Affect: when *Cognition* (reasoning) was high with respect to *Affect* (emotion) *Redemption* tended to be lower.

4. Network properties as measured by the metric *% Giant Component* indicated that there is a relationship between linguistic markers in terms of *Anger, Sadness, Cognition* and *Redemption*, and connectedness with a network. Thus, increased *Anger* relative to *Sadness* was, in the Ferguson data, accompanied by increased connectedness. Likewise for *Redemption* and (inversely) for *Cognition* relative to *Affect*. Language patterns appear to correlate with measures of network centrality. If we assume that increased centrality also increases social influence, we have additional evidence that contagion processes may operate within some social networks to increase the level of conditions conductive to risk-seeking.

Research Recommendations

The research reported here was somewhat constrained in what it could do given the conditions under which data were collected. Ideally, the research would have had access to historical social media (in this case Twitter) that could be used to create databases that are equivalent in scope and depth across a range of social events. We were not able to do that due to constraints of the public API. In addition, were we able, we would have pursued understanding more about the relationship between linguistic markers, network characteristics (centrality) and individual users. Again, we were constrained from those kinds of analyses by the public nature of our data access.

Nonetheless, the results of the present study are promising in that they provide an indication that a combination of linguistic and network properties analysis can provide an indication of situations conductive to disruptive behavior. Additional research along the lines of the Redemption concept could be undertaken. For example, other linguistic dimensions could be developed that reflect an action orientation to social situations.

Second, a broader range of social events could be studied given historical data access. This could include examining social events when the social environment was quickly changing. Examples of such events could include:

- Political events disrupted by demonstrations;
- Terrorist attack occurrences;
- Protestors either attacked or are successful in their quest;
- Disruptive events occurring within a widely-watched sporting event, such as a Super Bowl or World Cup final game.

Another research approach that builds on the present research would be to combine the methods developed here with other more detailed information concerning a particular set of disruptive events that provides a track of behaviors in the ground environment. For example, in the case of Ferguson, additional information in the form of police records, 911 call logs and the like would give an indication of occurrences on the ground that would be correlated with Twitter-relate metrics such as those research here.

On a larger scale and with a larger number of events, differences between events in terms of social responses (e.g., violence, disruption) could be correlated with base rate information about communities, etc., in terms of trends in violence crime and other social indicators that may, in combination with metrics applied to social media, give a clear and more predictive picture of when and where social disruption is more likely to occur. The application of Bayesian modeling techniques could be used to develop *a priori* distributions of social disruption based on background rates of crime and the like, with social media metrics providing Bayesian updating.

APPENDICES

Appendix A: Graph Creation and Analytics

Graph Creation and Analytics

1. Graph Generation

The purpose of graph generation is to extract and map the embedded relationships in the tweets modeled as a *social interaction graph*. The graph is used to find graph metrics that will be used to investigate network structure.

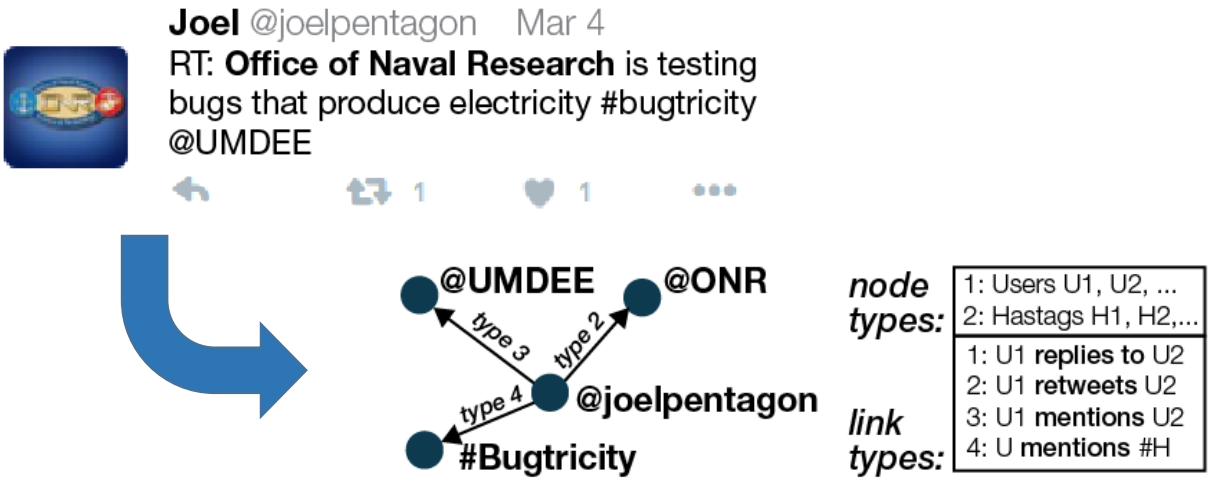

Figure 1: Process of creating a graph from a tweet.

The process of creating a graph is depicted in Figure 1. We first extract the users and hashtags from the tweets, determine whether it is a regular tweet, retweet, or a reply, and correspondingly create nodes and edges in the graph based on the tweets and their relationships to other tweets. Formally, let $G = (V, E)$ be the social interaction graph where V is the set of nodes ($|V| = n$) and E is the set of edges. A user or hashtag $v \in V$ if v is the user tweeting the tweet, or occurs in the tweet. Note that users always start with @ and hashtags always start with #. An edge, i.e., a two-tuple of nodes, $(u, v) \in E$ if and only if (i) a user u replies to v, (ii) a user u retweets another user v, (iii) a user u mentions another user v, or (iv) a user u mentions a hashtag v. While there is an implied directionality, for our purposes, we treat the graph G as undirected.

2. Graph Metrics

In this section, we discuss how we compute the graph metrics of the social interaction graph of G.

Degree Centrality: Nodes which are more central in the network than other nodes have higher centrality. A node's importance in a network has been related to its centrality (Ghosh and Lerman, 2010). Each node's centrality is measured by how many other nodes are adjacent to it. We compute the degree centrality of a node $v_i \in V$ as

$$C_D(v_i) = d_i = \sum_j A_{ij},$$

where A is the binary adjacency matrix of $n \times n$ such that $A_{ij} = 1$ if and only if nodes $i, j \in V$ are adjacent, i.e., they are neighbors, and 0 otherwise. The normalized degree centrality is then defined as

$$C'_D(v_i) = \frac{d_i}{n-1}.$$

Note that $0 \leq C'_D(v_i) \leq 1$.

PageRank Centrality: PageRank Centrality (C_P) is another indicator of node centrality. It considers both (1) the number of input links to any node and (2) the importance of each of the nodes that vote for it. While counting input links is straightforward (popularity of a node), weighting the importance of nodes that vote for the measured node is affected by a number of factors. One of these is the popularity of the voting node: a node that has many input links itself counts for more as a voting node that another node that no other nodes link to. A second is the relative number of nodes a voting node links to: if a voting node links to many nodes, these count less than if the voting node linked to only a few nodes: this shows more discriminability of the voting node. PageRank centrality metrics present a probability distribution of the likelihood that a random graph traversal on the graph will arrive at a particular node. This traversal is affected both by the popularity of the targeted node and the importance of the voting nodes and its discriminability. Together, the PageRank centrality roughly represents how central (or important) a targeted node is to the graph. There are different implementations of PageRank. We used an iterative one with the damping factor α. The PageRank centrality of a node is then represented as

$$C_P(v_i) = \frac{1-\alpha}{n} + \alpha \sum_{v_j \in M(v_i)} \frac{C_P(v_j)}{L(v_j)},$$

where $M(v_i)$ is the set of links that direct to v_i, i.e., inbound links, and $L(v_i)$ is the set of links that are directed from v_i, i.e., outbound links. Note that, for undirected graphs, $M(v_i) = L(v_i)$.

Average (Global) Clustering Coefficient: Networks differ in the degree to which their nodes cluster together. Networks in which two connected nodes are also connected to the same third node exhibit more clustering than a network in which two connected nodes are not connected to the same third node. This is computed based on triplets of nodes. A triplet consists of three nodes that are connected by either two (open triplet) or three (closed triplet) undirected ties. A triangle consists of three closed triplets, one centered on each of the nodes. Then the average (global) clustering coefficient is defined as

$$GCC = \frac{3 \times (\#\ of\ triangles)}{\#\ of\ connected\ triplets\ of\ vertices}.$$

Transitivity: The transitivity is closely related to the average clustering coefficient however we only count the triads, i.e., two edges with a shared vertex. Hence we define transitivity as

$$TR = \frac{3 \times (\#\ of\ triangles)}{\#\ of\ triads}.$$

Note that both the degree centrality and the PageRank centrality creates series over V. Hence we can denote degree centrality series as $\mathbf{C_D'} = \{C_D'(v_i) | v_i \in V\}$ and $\mathbf{C_P} = \{C_P(v_i) | v_i \in V\}$**,** and define first and second order statistics on them. In particular, we define the mean of the degree centrality as $\overline{\mathbf{C_D'}} = (1/n) \sum_{v_i \in V} C_D'(v_i)$, the maximum of the degree centrality as $\mu(\mathbf{C_D'}) = \max_{v_i \in V} C_D'(v_i)$, and the variance of the degree centrality as $\sigma^2(\mathbf{C_D'}) = \left(\frac{1}{n}\right) \sum_{v_i \in V} (C_D'(v_i) - \overline{\mathbf{C_D'}})^2$. The mean, maximum, and the variance of the PageRank centrality is also similarly defined.

Giant Component: A giant component refers to the % of nodes in the network that are connected to the largest subgraph (set of connected nodes) from the original network. As the network is more connected, the giant component score increases. The giant component of the graph G, namely $G_0 = (V_0, E_0) \subseteq G$, is defined as the largest connected subgraph of G. Then the percentage of the giant component size is defined as $100 * |V_0|/n$.

In our analysis, for each collection of tweets and based on the graph created on this collection, we report (i) the number of retweets, (ii) the number of replies, (iii) the number of nodes (n), (iv) the number of tweets per hour, (v) the number of edges ($|E|$), (vi) edges per node ($|E|/n$), (vii) average degree centrality ($\overline{\mathbf{C_D'}}$), (viii) maximum degree centrality **(**$\mu(\mathbf{C_D'})$), (ix) variance of the degree centrality ($\sigma^2(\mathbf{C_D'})$), (x) average PageRank centrality ($\overline{\mathbf{C_P'}}$), (xi) maximum PageRank centrality ($\mu(\mathbf{C_D'})$), (xii) variance of the PageRank centrality ($\sigma^2(\mathbf{C_P'})$), (xiii) average clustering coefficient (GCC), (xiv) transitivity (TR), (xv) percentage of isolated nodes ($100 * \omega/n$), and (xvi) the percentage of the giant component ($100 * |V_0|/n$).

Bot Detection: Understanding the social network structure of tweets is complicated by the existence of tweets that are created not by humans but by computer programs ("bots"). Bots can be a considerable amount of the traffic on a network defined by a common hashtag. Analysis of social networks that use twitter data need to sometimes (a) analyze datasets that have bots removed, as the question of interest only relates to human participants in the network; and (b) include all tweets, regardless of human or computer origin, as the analysis must examine the tweets that comprise the entire network.

Before discussing how we distinguished human tweets from bot tweets, we first will clarify characteristics of tweet datasets that will be analyzed for bots. A dataset can be all tweets collected during a time segment that contain a specific hashtag set, that use a specific language, or are sent from a known area. In this tweet dataset, we have tweets from *M* different usernames. Each username can have *1* to *N* number of tweets.

More formally, we assume a user submits a set $\boldsymbol{Q}$ queries on the variables of interest, e.g., keyword, language, area, time period, and namely $\boldsymbol{a_i, i = 1, 2, \ldots, Q}$, at time $\boldsymbol{t}$. We then *slice* the database with the query $q_t = \bigcap_{i=1}^{Q} q_t(a_i)$ and retrieve $\boldsymbol{N_{q_t}}$ tweets $\bigcup_{i=1}^{N_{q_t}} \tau_{\omega_i}$ uniquely identified by the indices $\boldsymbol{\omega_i}$.

We can also associate the tweets with the users that tweeted. Hence the collection of tweets $\bigcup_{i=1}^{N_{q_t}} \tau_{\omega_i}$ can be refactored as $\bigcup_{i=1}^{M_{q_t}} \bigcup_{j=1}^{N_i} \tau_{\omega_j}(\theta_i)$, where M_{q_t} is the number of users in this collection and each user θ_i has N_i tweets in this collection.

We then applied a bot detection scoring system to every username. The scoring system used the set of features described in Table X. Table X describes the factors examined to classify usernames as bots, and how these features are associated with the likelihood that a username is a bot and not a human.

Factors associated with likelihood a username is a bot	**Reasons for associating a factor with increased likelihood that a user is a bot**
Very high tweet rate	If a user sends out tweets faster than a human can create a tweet, the user is likely a program.
Higher follower/friend ratio	A program is not likely to be a friend, though other bots can follow it.
Post similar tweets to everyone	A program is often used to transmit a message, so it sends the same message to everyone.
User account were created more recently	Human creators of bot may make a new bot when they wish to transmit a new message
Retweet more	Bots are often used to transmit a message, so retweet others' tweets.
Post many tweets	To transmit a message, send out the message more.
Periodically send tweets	Bots generally send messages at regular intervals that do not follow circadian rhythms of humans.

After applying a bot score algorithm to each username in the current dataset, each username has a "bot score" that presents the likelihood that its tweets were generated by a bot. Bot scores are used to create ranking of which usernames are bots.